Good generalizability of a prediction rule for prediction of persistent shoulder pain in the short term.

Kuijpers, T.; van der Heijden, G.J.; Vergouwe, Y.; Twisk, J.W.R.; Boeke, AJ; Bouter, L.M.; van der Windt, D.A.W.M.

published in
Journal of Clinical Epidemiology
2007

DOI (link to publisher)
10.1016/j.jclinepi.2006.11.015

document version
Publisher's PDF, also known as Version of record

ELSEVIER

Journal of Clinical Epidemiology 60 (2007) 947−953

Journal of Clinical Epidemiology

Good generalizability of a prediction rule for prediction of persistent shoulder pain in the short term

Ton Kuijpers[a,b,c,*], Geert J.M.G. van der Heijden[d], Yvonne Vergouwe[d], Jos W.R. Twisk[e], A. Joan P. Boeke[a], Lex M. Bouter[a], Daniëlle A.W.M. van der Windt[a,f]

[a]*Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands*
[b]*Department of Allied Health Care Research, Amsterdam School of Allied Health Care Education, The Netherlands*
[c]*Care and Public Health Research Institute, Maastricht University, The Netherlands*
[d]*Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands*
[e]*Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands*
[f]*Primary Care Musculoskeletal Research Centre, Keele University, Keele Staffordshire, UK*

Accepted 14 November 2006

Abstract

Objective: To evaluate the generalizability of recently developed clinical prediction rules for the prognosis of shoulder pain in general practice.

Study Design and Setting: A large research program, consisting of a prognostic cohort study and three randomized controlled trials with 6 months follow-up, was carried out in The Netherlands. The clinical prediction rules were derived from the results of the prognostic cohort study ($n = 587$). The main outcome measure was persistent symptoms at 6 weeks or 6 months. The control groups of the trials who received usual care were merged ($n = 212$), and used to validate the prediction rules by studying calibration and discrimination.

Results: The prediction rule for short-term outcome showed reasonable calibration and discriminative ability in this validation cohort. The area under the receiver operating characteristic curve (AUC) was 0.72 compared to 0.74 in the derivation cohort. The prediction rule for long-term outcome performed less well. Discriminative ability (AUC) decreased to 0.56 in the validation cohort compared to 0.67 in the derivation cohort.

Conclusion: The prediction rule for the short-term (6 weeks) prognosis showed good generalizability. The prediction rule for the long-term prognosis showed poor generalizability.

Keywords: Shoulder; Musculoskeletal diseases; Prognosis; Cohort studies; Validation studies; Family practice

1. Introduction

Shoulder pain is common with a 1-year prevalence ranging between 5% and 47% [1−5]. The prevalence in the general population in The Netherlands has recently been estimated at 17% [6]. The annual incidence of consultation for a new episode of shoulder pain in Dutch general practice ranges between 12 and 25/1000/yr [3,6−8]. Shoulder pain has an unfavorable outcome in many patients. About 40−50% of all patients who present with a new episode of shoulder pain in primary care report persistent symptoms after 6−12 months [9−11].

We developed clinical prediction rules consisting of a limited number of (easily measurable) prognostic factors to predict the risk of persistent shoulder symptoms at the short (6 weeks) and long term (6 months). Such information may also support decisions regarding treatment and referral of patients. The performance (i.e., calibration and discrimination) of the prediction rules was evaluated in the development study [12]. Calibration refers to what extent the observed frequencies agree with the predicted probabilities. Discrimination refers to the ability to distinguish between a patient with persistent symptoms and a patient without persistent symptoms.

Before considering implementation of the prediction rules in clinical practice their generalizability needs to be tested [13−15]. Generalizability refers to the performance in patients drawn from a different but comparable population [13]. Our objective was therefore to evaluate the performance of our clinical prediction rules for the prognosis of shoulder pain in a different population of patients with shoulder pain in primary care.

* Corresponding author. Tel.: +31-20-652-1172; fax: +31-20-652-1141.
E-mail address: a.c.kuijpers@hva.nl (T. Kuijpers).

0895-4356/07/$ − see front matter
doi: 10.1016/j.jclinepi.2006.11.015

2. Methods

2.1. Dutch shoulder study

The Dutch Shoulder Study (DSS) is a comprehensive cohort study, carried out between January 2000 and May 2005. The DSS consists of one prognostic cohort study and three randomized controlled trials, which were carried out alongside each other. Between January 2001 and June 2003, 103 general practitioners (GPs) recruited patients at first consultation for a new episode of shoulder complaints in three geographic areas in The Netherlands (Amsterdam, Groningen, and Maastricht). All patients in the DSS had to meet the same general inclusion criteria, and specific additional inclusion criteria if eligible for a trial (Table 1). For the prognostic cohort study no additional inclusion criteria were specified. Data of the prognostic cohort study were used to derive the prediction rules. Data of the control groups of the three trials were used to study the generalizability of the rules.

The Groningen Manipulation Study [16,17] studies the effectiveness of manipulative therapy for the shoulder girdle in addition to usual care. In two other trials, a Graded Exercise Therapy [18] and an Education and Activation Program [19], respectively, were studied. Patients in the control groups of the trials received usual care, similar to the patients in the cohort study.

Baseline and follow-up assessments for all patients in the DSS were identical. The outcome was measured by postal questionnaires at 6 weeks, 3, and 6 months. The primary outcome measure was ''patient perceived recovery'' measured on an 8-point scale (1–3 = very much/much/slight deterioration; 4 = no change; 5–7 = slight/much/very much improvement; 8 = complete recovery) [17,20]. Persistent symptoms were defined as any deterioration, no change, slight or much improvement. Secondary outcome measures were shoulder disability, measured with the 16-item shoulder disability questionnaire (0–100) [21], pain (0–10 numeric rating scale) [20], and severity of the main complaint (0–10 numeric rating scale) [22].

Table 1
Selection criteria for the DSS

General inclusion criteria
Patients 18 yr of age or older
Not consulted GP or received any form of treatment for the afflicted shoulder in the preceding 3 mo
Sufficient knowledge of the Dutch language
Specific inclusion criteria trials
Groningen Manipulation Study (GMO)
Dysfunction of the cervicothoracic spine and adjacent ribs with accompanying pain or restricted movement
Graded Exercise Therapy Study (GET)
Duration of complaints >3 mo
Education and Activation Program (EAP)
Duration of complaints <3 mo
Exclusion criteria
Severe physical or psychological conditions (i.e., fractures or luxation in the shoulder region; rheumatic disease; neoplasm; neurological or vascular disorders; dementia)

2.2. Management of shoulder pain

All participants in the cohort study and the trial participants randomized to the control groups, received standardized treatment according to the 1999 version of the Dutch guidelines for shoulder disorders issued by the Dutch College of General Practitioners [23,24]. The guidelines recommend giving information on the prognosis of shoulder pain, advice regarding provoking activities, and stepwise treatment consisting of paracetamol, nonsteroidal antiinflammatory drugs (NSAIDs), corticosteroid injection, or referral for physiotherapy. The GP made the decision regarding the content of treatment based on duration and severity of pain and disability. The participating GPs were educated and trained to apply treatment according to this guideline.

2.3. Prediction rules

The prediction rules for persistent symptoms (yes/no) after 6 weeks and 6 months were developed using information from the 587 patients of the derivation cohort. Sociodemographic variables, disease characteristics (i.e., pain intensity, disability, duration of complaints, onset, comorbidity), physical workload, psychosocial factors, and results of a physical examination were used to derive the prediction rules. We tested the internal validity with bootstrapping techniques [13]. The calibration of the prediction rules was adequate. The discriminative ability was satisfactory with area under the receiver operating characteristic (ROC) curve (AUC) of 0.74 (95% confidence interval [CI] 0.70, 0.79) at 6 weeks and 0.67 (95% CI 0.63, 0.71) at 6 months. Fig. 1 presents the prediction rules as score charts, development of which has been described in detail elsewhere [12].

2.4. Analysis

We compared the baseline characteristics of the derivation and validation cohort to assess differences of $\geqslant$10%. This cutoff point was used as a rule of the thumb to explore whether there are differences between the cohorts that may explain a reduction in the performance of the prediction rule when tested in the validation cohort. The performance of the prediction rules was tested in the validation cohort by evaluating their calibration and discrimination. Calibration was assessed by plotting the predicted probabilities of persistent symptoms according to the prediction rule, against the observed frequencies. For this, patients were grouped into quintiles according to their predicted probability of persistent symptoms. The prevalence of the endpoint within each quintile equals the observed frequency. A more formal

Score chart for prediction of persistent shoulder symptoms at 6 weeks

Duration of complaints		
<6 weeks	0	…
6-12 weeks	7	…
>3 months	11	…
Gradual onset	7	…
Psychological complaints	10	…
Repetitive movements	8	…
Shoulder pain (0-10)	*score*	…
Neck pain score at physical examination (0-18)	*score*	…
		+
Total score		…

Total score	Risk
≤2	20% - 30%
3 – 7	30% - 40%
8 – 11	40% - 50%
12 – 16	50% - 60%
17– 21	60% - 70%
22 – 27	70% - 80%
28 – 36	80% - 90%
≥37	90% - 100%

The predicted probability of persistent symptoms at 6 weeks was determined by P=1/[1+ exp – (–1.19 + 0.64 × duration of complaints 6-12 weeks + 0.95 × duration of complaints >3 months + 0.59 × gradual onset + 0.85 × concomitant psychological complaints + 0.68 × repetitive movements + 0.13 × shoulder pain + 0.09 × neck pain score at physical examination)].

Score chart for prediction of persistent shoulder symptoms at 6 months

Duration of complaints		
<6 weeks	0	…
6-12 weeks	9	…
>3 months	17	…
Gradual onset	10	…
Concomitant low back pain	13	…
Shoulder pain (0-10)	*score* × 2	…
Shoulder pain score at physical examination (0-18)	*score*	…
		+
Total score		…

Total score	Risk
≤1	10% - 20%
2 – 16	20% - 30%
17 – 28	30% - 40%
29 – 39	40% - 50%
40 – 49	50% - 60%
50 – 61	60% - 70%
≥62	70% - 100%

The predicted probability of persistent symptoms at 6 months was determined by P=1/[1+ exp – (–1.48 + 0.34 × duration of complaints 6-12 weeks + 0.64 × duration of complaints >3 months + 0.37 × gradual onset + 0.50 × concomitant low back pain + 0.08 × shoulder pain + 0.04 × shoulder pain score at physical examination)].

Instruction
If a predictor is scored positively, the given weight needs to be filled in. Subsequently the scores are added to calculate the 'Total score'. From the table next to the score chart the risk (%) of persistent symptoms for an individual patient can be determined.

Fig. 1. Prognostic score charts for prediction of persistent symptoms at 6 weeks and 6 months.

indication of calibration can be obtained by fitting a logistic regression model with the logodds of the predicted risks as only covariate. This model has an intercept and a slope. If predicted risks and observed frequencies are in agreement, the intercept is equal to 0 and the slope equal to 1.

The AUC was used to assess the discriminative ability of the model. An AUC of 0.5 indicates no discrimination above chance, whereas an AUC of 1.0 indicates perfect discrimination. Because the discriminative ability of a rule is related to the homogeneity of the sample in which the rule is applied, we estimated the maximum attainable AUC. Using the predicted risks of the patients in the validation cohort, outcomes were generated with Monte Carlo Simulation [25,26]. This mimics the situation that the predictions were correct, that is, the model is perfectly calibrated. The AUC that was estimated for the predicted risks and generated outcomes was considered the maximum attainable AUC for the validation sample.

Furthermore, to gain insight into the performance of our prediction rules, we estimated the multivariable logistic regression coefficients for each of the predictors of our prediction rule in the validation cohort. This analysis shows which of the different elements of the rule are the strongest predictors of persistent shoulder pain in the validation cohort.

3. Results

3.1. Study population

Table 2 presents the baseline characteristics of the derivation cohort and validation cohort. Patients in the

validation cohort clearly showed a longer duration of complaints at baseline (13% less often complaints between 0 and 6 weeks, and 14% more often complaints >3 months), and reported 10% more neck complaints in the past in comparison with the derivation cohort.

3.2. Course of symptoms

Table 2 shows that more patients in the validation cohort reported persistent symptoms after 6 weeks (89% vs. 70%) and 6 months (69% vs. 46%), compared to the derivation cohort. Nevertheless, patients in the derivation and validation cohort reported similar levels of pain and disability at the different time points. Likewise, patients reporting recovery in the validation cohort and in the derivation cohort showed similar low levels of pain (<1 point), disability (<13 points), and severity of the main complaint (<1 point).

3.3. Management of shoulder pain

At baseline most patients in the derivation cohort ($n = 423$, 72%) received a wait and see policy, paracetamol, or NSAIDs. Furthermore, 68 patients (12%) received an injection with corticosteroid, 58 patients (10%) were referred for physiotherapy, and 28 patients (6%) received other therapies. In the validation cohort, 141 patients (83%) received a wait and see policy, paracetamol, or NSAIDs; 9 patients (5%) received an injection; 11 patients (7%) were referred for physiotherapy; and 8 patients (5%) received other therapies.

3.4. Performance

Fig. 2 shows the calibration of the predictions. For 6 weeks, the plotted points were rather close to the 45° line, although most predictions slightly underestimated the

Table 2
Description of baseline characteristics and outcome measures at 6 wk and 6 mo of patients with shoulder pain in the derivation ($n = 587$) and validation cohort ($n = 212$)

	Derivation		Validation	
Baseline characteristics				
Demographic				
Age (yr); mean (SD)	51 (14)		51 (12)	
Gender: male; *n* (%)	292 (50)		92 (44)	
Disease characteristics				
Duration of complaints; *n* (%)				
0–6 wk[a]	205 (35)		46 (22)	
7–12 wk	139 (24)		49 (23)	
>3 mo	242 (41)		115 (55)	
Gradual onset (vs. acute); *n* (%)	363 (62)		144 (69)	
Precipitating cause; *n* (%)				
Strain/overuse: usual activities	138 (24)		58 (28)	
Shoulder complaints in the past; *n* (%)	348 (62)		136 (65)	
Neck complaints in the past; *n* (%)	296 (51)		128 (61)	
Comorbid psychological complaints; *n* (%)	55 (9)		20 (10)	
Concomitant musculoskeletal complaints; *n* (%)				
Neck/high back	209 (36)		85 (41)	
Low back pain	139 (24)		59 (28)	
Upper extremity	174 (30)		76 (36)	
Shoulder pain (0–10); mean (SD)	4.8 (2.3)		5.3 (2.2)	
Shoulder disability (0–100); mean (SD)	60 (24)		62 (24)	
Physical examination				
Shoulder pain score (0–18); median (IQR)	4 (2–4)		7 (4–7)	
Neck pain score (0–18); median (IQR)	0 (0–0)		2 (0–2)	
Physical factors				
Dynamic physical workload (0–5); median (IQR)	1 (1–2)		1 (0–1)	
Repetitive movements; *n* (%)	384 (65)		151 (73)	
Outcome measures	6 wk	6 mo	6 wk	6 mo
Persistent symptoms; *n* (%)	340 (70)	249 (46)	161 (89)	125 (69)
Pain[a] (0–10); mean (SD)	4.3 (2.1)	4.1(2.3)	4.3 (2.1)	4.0 (2.0)
Shoulder disability[a] (0–100); mean (SD)	53.0 (25.5)	52.2 (26.7)	56.0 (25.6)	54.4 (27.2)
Severity of main complaint[a] (0–10); mean (SD)	4.8 (2.6)	5.0 (2.8)	4.9 (2.5)	5.6 (2.6)

SD, standard deviation; IQR, interquartile range.
[a] Mean and SD presented for group reporting persistent symptoms.

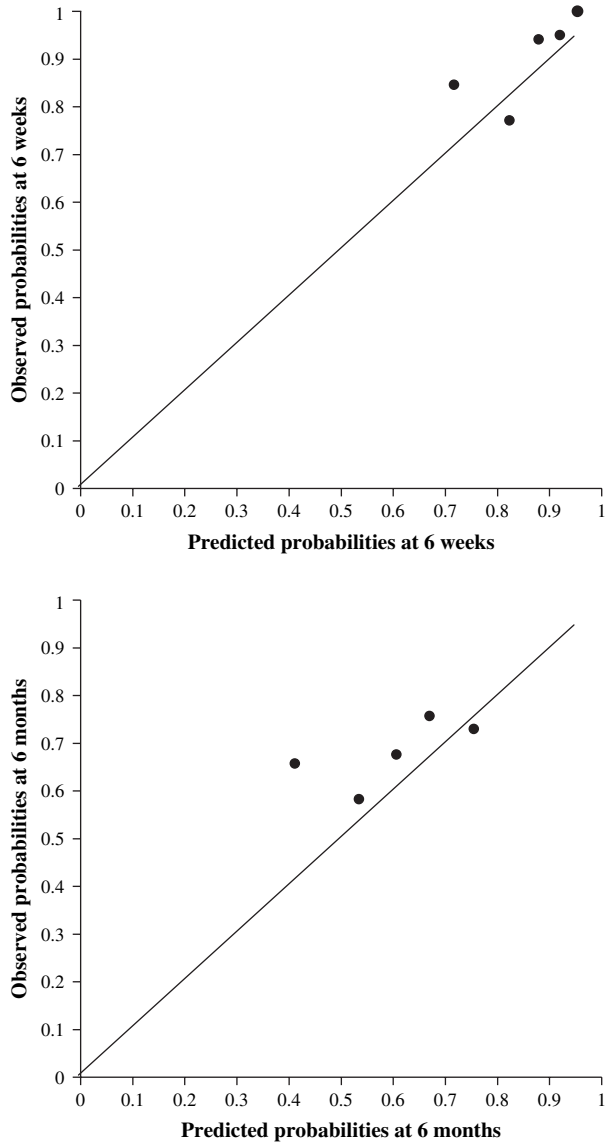

Fig. 2. Calibration plots showing the observed probabilities vs. the predicted probabilities for persistent symptoms at 6 weeks ($n = 175$) and 6 months ($n = 180$) in the validation cohort.

observed probabilities. This is in agreement with an intercept of 0.44 in a model, using the logodds of the predicted probabilities as only covariate. The intercept above 0 confirms that the predicted probabilities were generally too low. The slope of the model was 1.1, which is close to 1.

For 6 months, the plotted points were further away from the 45° line, demonstrating a rather poor calibration of the model. This is confirmed by an intercept of 0.39 (not close to 0) and a slope of 0.43 (not close to 1).

The discriminative ability (AUCs) of the prediction rules was 0.72 (95% CI 0.63, 0.82) at 6 weeks, and 0.57 (95% CI 0.48, 0.66) at 6 months. The Monte Carlo Simulation showed a maximum attainable AUC of 0.70 at 6 weeks, and 0.64 at 6 months.

Table 3 shows the estimates of the regression coefficients for each predictor when the rule is applied to the validation cohort. A duration of complaints >3 months and repetitive movements were strong predictors of persistent symptoms at 6 weeks. For long-term outcomes, a duration >3 months and a gradual onset seemed strongly related with persistent symptoms at 6 months.

4. Discussion

The performance of the prediction rule for the short-term (6 weeks) prognosis of shoulder pain in the validation cohort was satisfactory. There are no generally accepted cutoffs for the AUC, although a value of 0.80 or higher is sometimes used as a rule of the thumb to indicate good performance. Calibration and discriminative ability were reasonable, and similar to that found in the derivation cohort (AUC = 0.74 in derivation cohort, and AUC = 0.72 in validation cohort). The prediction rule for long-term (6 months) outcome showed poor calibration and discrimination. The AUC decreased from 0.67 in the derivation cohort to 0.57 in the validation cohort, not much more than a flip of a coin (AUC = 0.50), which means that the performance of the long-term prediction rule was disappointing.

RCTs usually include more homogeneous patient populations. This may have affected the performances of the prediction rule in the validation cohort, but at the same time demonstrates the applicability of the rule in a different setting and slightly different population. The patients in the validation cohort differed from the derivation cohort regarding several aspects. In general, the shoulder complaints from the patients in the validation cohort were more severe (Table 3). They showed a longer duration of symptoms, which is an important predictor of outcome [12], and reported more neck complaints in the past. This may indicate that we have tested the performance of the prediction rule in patients who, on average, were more advanced in their disease process or who may have had a somewhat different type of shoulder problem (which has been described as spectrum transportability [13]). More severe complaints at baseline may have resulted in more frequent reports of persistent symptoms (Table 3). Another possible explanation for the higher occurrence of persistent symptoms could be that the validation cohort more often received a wait and see policy (83% vs. 72%) and were less frequently treated with local infiltration of a corticosteroid (5% vs. 12%).

It has been reported that a validation cohort should minimally consist of 100 events [26]. The event rate was sufficient for the prediction rule for 6 months. However, the performance of the rule for 6 weeks was better than the rule for 6 months. This may indicate that the lack of sufficient events may be of less importance than the relative weak associations between baseline characteristics and outcome.

Differences in prognosis between the derivation and validation cohort may have substantially altered the calibration of our prediction rules in the validation cohort, especially

Table 3
Prediction rules for persistent shoulder symptoms at 6 wk and 6 mo after first consultation

Predictor	Scale	Derivation		Validation	
		β	95% CI	β	95% CI
6 weeks					
Duration of complaints					
0 to 6 wk[a]		—		—	
7 to 12 wk	(yes/no)	0.64	0.1 to 1.2	0.42	−0.9 to 1.7
>3 mo	(yes/no)	0.95	0.4 to 1.5	1.67	0.2 to 3.1
Gradual onset	(yes/no)	0.59	0.1 to 1.1	0.36	−0.7 to 1.4
Concomitant psychological complaints	(yes/no)	0.85	−0.1 to 1.9	0.74	−1.5 to 3.0
Repetitive movements	(yes/no)	0.68	0.2 to 1.1	1.05	−0.1 to 2.2
Shoulder pain	(0 to 10)	0.13	0.0 to 0.2	0.06	−0.2 to 0.3
Neck pain score at physical examination	(0 to 18)	0.09	0.0 to 1.0	0.03	−0.1 to 0.2
6 months					
Duration of complaints					
0 to 6 wk[a]		—		—	
7 to 12 wk	(yes/no)	0.34	−0.1 to 0.8	−0.10	−1.3 to 1.1
>3 mo	(yes/no)	0.64	0.2 to 0.1	1.24	−0.1 to 2.6
Gradual onset	(yes/no)	0.37	0.0 to 0.6	0.70	−0.3 to 1.7
Concomitant low back pain	(yes/no)	0.50	0.1 to 0.9	0.01	−1.1 to 1.2
Shoulder pain	(0 to 10)	0.08	0.0 to 0.2	0.04	−0.2 to 0.3
Shoulder pain score at physical examination	(0 to 18)	0.04	0.0 to 0.1	0.09	−0.0 to 0.2

The β values are derived from a multiple logistic regression analysis. The β values for the validation cohort are derived from the results of a multiple logistic regression analysis conducted with the predictors from the derivation cohort.

[a] Reference category.

for the long term. The reason for this is that statistical models are calibrated to the overall outcome prevalence. For a substantial part, this prevalence is determined by the characteristics of the patient population. As long as the overall prevalence is explained by predictors which are included in the prediction rule, the model will still be well calibrated. This may have resulted in a reasonable calibration of our short-term prediction rule, despite differences at baseline between the derivation and validation cohort regarding important predictors, that is, duration of complaints and repetitive movement. The poor performance of the prediction rule for long-term outcomes may be explained by patient characteristics which are not documented in this study, but yet strongly influenced outcome in the validation cohort.

The maximum attainable AUC for the short term of 0.70 strengthens our findings of adequate discriminative ability of this rule. Calibration was reasonable, although an intercept of 0.44 showed that the risk of persistent shoulder pain was generally underestimated (see also Fig. 2). For the long term, a substantially lower maximum attainable AUC of 0.64 differed considerably from the achieved AUC of 0.57. This indicated a model with poor discriminative ability. Regression coefficients were generally too high, and insufficient shrinking of the regression coefficients had been achieved in the development stage of the prediction rule. Justice et al. [13] stated that perhaps the most difficult test of discrimination occurs when the spectrum of a disease narrows from both sides; that is, the test sample includes many patients who have an illness of intermediate severity and very few who are either severely ill or not very ill at all. This could partly explain the poor performance of our long-term prediction rule as most observed probabilities of persistent symptoms were distributed between 0.5 and 0.7 (Fig. 2). This reflects a homogeneous population resulting in a low maximum attainable AUC of 0.64.

We developed prediction rules to predict the prognosis of shoulder pain in general practice. Most elements of the prediction rules were derived from a questionnaire, filled out by the patient. Our prediction rule might be applicable in primary care populations elsewhere, provided that management of shoulder pain is largely comparable. However, given the fact that performances of prediction rules may vary across populations, their predictive validity should ideally be tested when implemented in other settings or countries. If these prediction rules would be implemented in daily practice, it is the GP who asks the questions and calculates the risk by using a score chart. Or in a more sophisticated way, enters the responses into a personal computer or personal digital assistant, which calculates the risk of persistent symptoms. So, future research should also evaluate the methodologic transportability of the prediction rules (i.e., performance when data are collected by using alternative methods [13]) in a new sample of patients. And perhaps most importantly, the clinical usefulness of these instruments should be established: can the prediction rules be helpful to the clinician when making decisions in the management of patients with shoulder pain, for example, whether or not to consider additional diagnostic testing, start a certain treatment or refer the patient to secondary care [15].

Duration has indeed been shown to be an important predictor in previous research [27]. Our prediction, however,

includes a few other predictors that may help to estimate prognosis in primary care patients with shoulder pain. The main objective of this study, however, was not only to identify prognostic indicators of outcome, but also to present prognostic information in a simple format that enables the computation of the risk of a poor outcome in individual patients. This is the first time that a prediction rule for shoulder pain has been developed and tested.

In conclusion, the prediction rule for the short-term (6 weeks) prognosis of shoulder pain in general practice showed adequate generalizability in the validation cohort. The long-term outcomes (6 months) seems difficult to make accurate predictions of persistent shoulder pain in this population.

Acknowledgment

This study was supported by a grant (No. 904-65-09) from The Netherlands Organization for Scientific Research (NWO), The Hague, The Netherlands.

References

[1] Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. Scand J Rheumatol 2004;33:73–81.

[2] Van der Heijden GJMG. Shoulder disorders: a state-of-the-art review. Baillieres Best Pract Res Clin Rheumatol 1999;13:287–309.

[3] Bot SD, Van der Waal JM, Terwee CB, Van der Windt DAWM, Schellevis FG, Bouter LM, et al. Incidence and prevalence of complaints of the neck and upper extremity in general practice. Ann Rheum Dis 2005;64:118–23.

[4] Bongers PM. The cost of shoulder pain at work. BMJ 2001;322:64–5.

[5] Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. Pain 2003;102:167–78.

[6] Van der Linden MW, Westert GP, De Bakker DH, Schellevis FG. Complaints and disorders in the population and general practice. [Klachten en aandoeningen in de bevolking en in de huisartsenpraktijk]. [in Dutch]. Utrecht/Bilthoven: NIVEL/RIVM; 2005.

[7] Van der Windt DAWM, Koes BWK, Jong BAD, Bouter LM. Shoulder disorders in General Practice: incidence, patients characteristics, and management. Ann Rheum Dis 1995;54:959–64.

[8] Okkes IM, Oskam SK, Lamberts H. Van klacht naar diagnose: Episodegegevens uit de huisartspraktijk. Bussum: Coutinho; 1998.

[9] Croft P, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. BMJ 1996;313:601–2.

[10] Van der Windt DAWM, Koes BWK, Boeke AJP, Devillé W, De Jong BA, Bouter LM. Shoulder disorders in general practice: prognostic indicators of outcome. Br J Gen Pract 1996;46:519–23.

[11] Winters JC, Sobel JS, Groenier KH, Arendzen JH, Meyboom-de Jong B. The long-term course of shoulder complaints: a prospective study in general practice. Rheumatology 1999;38:160–3.

[12] Kuijpers T, Van der Windt DAWM, Boeke AJP, Twisk J, Vergouwe Y, Bouter LM, Van der Heijden GJMG. Clinical prediction rules for the prognosis of shoulder pain in general practice. Pain 2006;120(3): 276–85.

[13] Justice AC, Covinsky KE, Berlin JA. Assessing the generalisability of prognostic information. Ann Intern Med 1999;130:515–24.

[14] Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453–73.

[15] Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? Semin Urol Oncol 2002;20:96–107.

[16] Bergman GJ, Winters JC, Van der Heijden GJMG, Postema K, Meyboom-de Jong B. Groningen Manipulation Study. The effect of manipulation of the structures of the shoulder girdle as additional treatment for symptom relief and for prevention of chronicity or recurrence of shoulder symptoms. Design of a randomized controlled trial within a comprehensive prognostic cohort study. J Manipulative Physiol Ther 2002;25:543–9.

[17] Bergman GJ, Winters JC, Groenier KH, Pool JJ, Meyboom-de Jong B, Postema K, et al. Manipulative therapy in addition to usual medical care for patients with shoulder dysfunction and pain: a randomized, controlled trial. Ann Intern Med 2004;141:432–9.

[18] Geraets JJ, Goossens ME, De Bruijn CP, Koke AJ, De Bie RA, Pelt RA, et al. A behavioural treatment for chronic shoulder complaints: concepts, development, and study design. Aust J Physiother 2004;50:33–8.

[19] De Bruijn CP, De Bie R, Geraets JJ, Goossens ME, Koke AL, Van den Heuvel HW, et al. Evaluation of an education and activation programme to prevent chronic shoulder complaints: design of an RCT [ISRCTN71777817]. BMC Fam Pract 2005;6:7.

[20] Van der Windt DAWM, Koes BWK, Deville W, Boeke AJ, De Jong BA, Bouter LM. Effectiveness of corticosteroid injections versus physiotherapy for treatment of painful stiff shoulder in primary care: randomised trial. BMJ 1998;317:1292–6.

[21] Van der Heijden GJMG, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. J Clin Epidemiol 2000;53:29–38.

[22] Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, et al. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22:144–8.

[23] The Dutch College of General Practitioners (NHG) Practice Guideline. Available at. http://nhg.artsennet.nl/upload/104/guidelines2/E08.htm. Accessed September 1, 2003.

[24] Winters JC, Jongh AC, van der Windt DA, Jonquiere M, Winter AF, van der Heijden GJ, et al. Practice guideline ''shoulder complaints''. [NHG-Standaard Schouderklachten]. [in Dutch]. Huisarts Wet 1999;42:222–31.

[25] Gilks WR, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in practice. New York: Chapman & Hall; 1996.

[26] Vergouwe Y, Steyerberg EW, Eijkemans M, Habbema J. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58: 475–83.

[27] Kuijpers T, Van der Windt DAWM, Van der Heijden GJGM, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. Pain 2004;109:420–31.